

Computational Concepts: Homework 1

Due: Sept 29 th 2005 (by mid night)

Purpose of this to familiarize you with bulk handling and download of data from ftp sites and building blast database and using fastacmd to extract sequence.

1. Create a directory **hw-1** in your home on amadeus (remember to cd into it)

2. Download sequence files from the "ftp" site:

ftp://arl.arizona.edu/pub/igert-hw1/

(you can open it in your web browser to see how it looks)

This is how you get data in bulk from providers like NCBI

(<ftp://ftp.ncbi.nih.gov/>)

3. To make things easy we can download using the program ncftp while you are in the hw-1 directory type:

ncftp ftp://arl.arizona.edu/pub/igert-hw1

you will get a prompt looking like

ncftp /pub/igert-hw1 >

If you type **ls** you can list all the files

You can download all the file by using the command:

get *bulk.tfa.gz

4. Once downloaded, uncompress and combine them into a single file called:

allseq.tfa

5. Build a BLAST database call its **hw-1** (hint: -n hw-1)

6. Blast the collection of sequence in the file **unknown.tfa** located in:

/home/student/2005/eeb/hw-1

i.e /home/student/2005/eeb/hw-1/unknown.tfa

7. Extract the matching subject sequence for the top hit for all the 4 unknown sequence and save them as **AZ-1.tfa** , **AZ-2.tfa**, **AZ-3.tfa** , **AZ-4.tfa** respectively (make sure they are in **hw-1** directory)

8. Go to the online homework submission site and answer the questions.